Esercizi di Calcolo delle Probabilità e Statistica Matematica

Lucio Demeio
Dipartimento di Ingegneria Industriale e Scienze Matematiche
Università Politecnica delle Marche

1. **Esercizio (1 marzo 2011 n. 2).** *La legge individuata dalla densità*

$$f(t) = \frac{\lambda}{2}\, e^{-\lambda\,|t|}$$

è detta Legge di Laplace di parametro λ.

(i) Mostrare che f è una densità di probabilità;

(ii) calcolare media e varianza della legge di Laplace;

(iii) Se X è una variabile aleatoria con legge di Laplace di parametro λ, quali sono le leggi delle variabili $\alpha\, X$ e $|X|$?

Soluzione.

(i) Dobbiamo mostrare che l'integrale di $f(t)$ su tutto $\mathbb{R}$ è 1.

$$\int_{\mathbb{R}} f(t)\, dt = \int_{\mathbb{R}} \frac{\lambda}{2}\, e^{-\lambda\,|t|}\, dt = 2\int_0^{+\infty} \frac{\lambda}{2}\, e^{-\lambda\, t}\, dt = \int_0^{+\infty} \lambda\, e^{-\lambda\, t}\, dt = 1$$

(ii)

$$E[X] = \int_{\mathbb{R}} t\, \frac{\lambda}{2}\, e^{-\lambda\,|t|}\, dt = 0$$

$$Var(X) = E[X^2] - E[X]^2 = E[X^2] = \int_{\mathbb{R}} t^2\, \frac{\lambda}{2}\, e^{-\lambda\,|t|}\, dt = 2\int_0^{+\infty} t^2\, \frac{\lambda}{2}\, e^{-\lambda\, t}\, dt =$$

$$= \int_0^{+\infty} t^2\, \lambda\, e^{-\lambda\, t}\, dt = \Big[-t^2\, e^{-\lambda\, t}\Big]_0^{+\infty} + \int_0^{+\infty} 2\, t\, e^{-\lambda\, t} =$$

$$= \left[-2\, t\, \frac{e^{-\lambda\, t}}{\lambda}\right]_0^{+\infty} + 2\int_0^{+\infty} \frac{e^{-\lambda\, t}}{\lambda}\, dt = \frac{2}{\lambda^2}$$

(iii) Calcoliamo preventivamente la funzione di ripartizione di X, $F_X(t)$:

$$F_X(t) = P(X \le t) = \int_{-\infty}^{t} \frac{\lambda}{2}\, e^{\lambda\, t'} dt' = \frac{1}{2}\, e^{\lambda\, t}, \qquad t < 0$$

$$= \int_{-\infty}^{0} \frac{\lambda}{2}\, e^{\lambda\, t} dt + \int_0^{t} \frac{\lambda}{2}\, e^{-\lambda\, t'} dt' = 1 - \frac{e^{-\lambda\, t}}{2}, \qquad t \ge 0$$

Calcoliamo la funzioni di ripartizione di $\alpha\, X$.

$$F_{\alpha\, X}(t) = P(\alpha\, X \le t) = P\left(X \le \frac{t}{\alpha}\right) = F_X\left(\frac{t}{\alpha}\right), \qquad \alpha > 0$$
$$F_{\alpha\, X}(t) = P\left(X \ge \frac{t}{\alpha}\right) = 1 - P\left(X < \frac{t}{\alpha}\right) = 1 - F_X\left(\frac{t}{\alpha}\right), \qquad \alpha < 0$$

e quindi

$$\begin{aligned} f_{\alpha\, X}(t) &= (F_{\alpha\, X})'(t) = \frac{1}{\alpha} f\left(\frac{t}{\alpha}\right), \qquad \alpha > 0 \\ &= -(F_{\alpha\, X})'(t) = -\frac{1}{\alpha} f\left(\frac{t}{\alpha}\right), \qquad \alpha < 0 \end{aligned}$$

ovvero

$$f_{\alpha\, X}(t) = \frac{1}{|\alpha|} f\left(\frac{t}{\alpha}\right), \qquad \forall\ \alpha$$

Veniamo ora a $|X|$. Intanto abbiamo ovviamente

$$F_{|X|}(t) = 0, \qquad t \le 0.$$

Per $t > 0$:

$$F_{|X|}(t) = P(|X| \le t) = P\left(-t \le X \le t\right) = F_X(t) - F_X(-t) = 1 - e^{-\lambda\, t}$$

2. **Esercizio (1 marzo 2011 n. 3).** *In una schedina del totocalcio a* 13 *partite i tre simboli* 1*,* X *e* 2 *compaiono con probabilità* 0.46*,* 0.28 *e* 0.26 *rispettivamente. Calcolare la probabilità che in una schedina*

   (i) il 2 *compaia più di* 3 *volte;*

   (ii) il simbolo X *non compaia mai.*

Soluzione.

(i) Sia Y la v.a. che indica il numero di volte in cui appare "2 in una schedina. Dobbiamo calcolare $P(Y > 3) = P(Y \ge 4)$. La legge di Y è binomiale di parametri $n = 13$ e $p = 0.26$. La risposta pertanto è

$$P(Y \ge 4) = \sum_{k=4}^{13} \binom{13}{k} (0.26)^k\, (1 - 0.26)^{13-k} = 0.45$$

(ii) Sia ora Z la v.a. che indica il numero di volte in cui appare "X in una schedina. Dobbiamo calcolare $P(Z = 0)$. La legge di Z è nuovamente binomiale di parametri $n = 13$ e $p = 0.28$. La risposta pertanto è

$$P(Z = 0) = \binom{13}{0} (0.28)^0\, (1 - 0.28)^{13} = 0.014$$

3. **Esercizio (13 luglio 2011 n. 1).** *In una data università, la frazione di professori ordinari, professori associati, ricercatori ed istruttori è rispettivamente del* 30, 40, 20 *e* 10 *percento. Di questi, rispettivamete il* 90, 80, 70 *e* 20 *percento detiene il titolo di dottorato (altri tempi ...). Qualè la probabilità che una persona con il dottorato, presa a caso, sia un ricercatore?*

Soluzione. Si scelga a caso un docente dell'università in questione. Sia O l'evento "l'individuo sia un professore ordinario, A "l'individuo sia un professore associato, R "l'individuo sia un ricercatore ed I "l'individuo sia un istruttore. Analogamente, sia D l'evento "l'individuo possiede il dottorato. I dati del problema indiano che $P(O) = 0.3$, $P(A) = 0.4$, $P(R) = 0.2$ e $P(I) = 0.1$. Inoltre $P(D|O) = 0.9$, $P(D|A) = 0.8$, $P(D|R) = 0.7$ e $P(D|I) = 0.2$. La domanda richiesta è $P(R|D) =?$. Usando Bayes, abbiamo

$$P(R|D) = \frac{P(D|R)\,P(R)}{P(D)}$$

Il valore di $P(D)$ va calcolato usando il teorema delle probabilità totali, in quanto gli eventi O, R, A, I costituiscono una partizione dello spazio campione. Abbiamo:

$$P(D) = P(D|O)\,P(O) + P(D|A)\,P(A) + P(D|R)\,P(R) + P(D|I)\,P(I) = 0.75$$

e quindi

$$P(R|D) = \frac{P(D|R)\,P(R)}{P(D)} = \frac{0.7\,0.2}{0.75} = 0.187$$

4. **Esercizio (13 luglio 2011 n. 2).** *Siano* X *ed* Y *due variabili aleatorie continue. La densità congiunta è data da*

$$f_{XY}(x,y) = \begin{cases} x\,y/2, & 0 < y < x < 2 \\ \\ 0, & \textit{altrimenti} \end{cases}$$

Calcolare:

(i) $E[1/(XY)]$;

(ii) $E[X^2]$;

(iii) $E[X]$.

Soluzione. La distribuzione congiunta è diversa da zero nel triangolo di vertici $(0,0)$ (l'origine), $(2,0)$ e $(2,2)$.

(i)

$$E[1/(XY)] = \int\int_{\mathbb{R}^2} \frac{1}{x\,y}\, f_{XY}(x,y)\,dx\,dy = \int_0^2 dx \int_0^x dy\, \frac{1}{x\,y}\,\frac{x\,y}{2} =$$
$$= \int_0^2 dx\,\frac{x}{2} = 1$$

(ii)

$$E[X^2] = \int\int_{\mathbb{R}^2} x^2\, f_{XY}(x,y)\,dx\,dy = \int_0^2 dx \int_0^x dy\, x^2\,\frac{x\,y}{2} =$$
$$= \int_0^2 dx\,\frac{x^5}{4} = \frac{8}{3}$$

(iii) $E[X]$.

$$E[X] = \int\int_{\mathbb{R}^2} x\, f_{XY}(x,y)\,dx\,dy = \int_0^2 dx \int_0^x dy\, x\,\frac{x\,y}{2} = \frac{8}{5}$$

5. **Esercizio (12 novembre 2011 n. 4).** *Sia X una variabile aleatoria di legge $N(0,1)$ e sia Y la variabile aleatoria data da $Y = X$ se $X \geq 0$ ed $Y = 0$ altrimenti. Determinare la funzione di ripartizione di Y e dire se Y ammette una densità continua.*

Soluzione. Calcoliamo la funzione di ripartizione $F_Y(t)$ di Y. Notando che Y non è mai negativa, abbiamo che

$$F_Y(t) = P(Y \leq t) = 0, \qquad t < 0$$
$$F_Y(t) = P(Y \leq t) = P(X \leq t) = F_X(t) = \Phi(t), \qquad t \geq 0$$

dove $\Phi(t)$ è la funzione di ripartizione della distribuzione normale standard. La funzione di ripartizione di Y è dunque nulla per $-\infty < t < 0$, in $t = 0$ ha una discontinuità a salto e vale $F_Y(0) = 1/2$, seguendo poi il ramo di $\Phi(t)$ per $t \geq 0$. È una variabile di tipo misto, in quanto $P(Y = 0) = P(X \leq 0) = 1/2$ e non ammette densità continua.

6. **Esercizio (20 gennaio 2011 n. 1).** *N studenti, tra cui Mario, condividono un alloggio ed ogni settimana ne vengono estratti a sorte $k < N$ per compiere le pulizie, senza escludere quelli che sono stati sorteggiati in precedenza.*

   (a) *Determinare la probabilità p che Mario venga sorteggiato in una data settimana;*

   (b) *determinare la probabilità che Mario non debba mai fare pulizie per m settimane;*

   (c) *supponiamo che Mario non abbia mai fatto pulizie nelle prime r settimane; qual'è la probabilità che le debba fare alla $(r+1)$-esima settimana?*

   (d) *qual'è la probabilità che Mario debba fare le pulizie almeno s volte nelle prime l settimane?*

Soluzione.

(a) La probabilità p si determina come rapporto del numero di casi favorevolu sul numero di casi possibili. I casi possibili sono le combinazioni di N elementi di classe k,

$$\binom{N}{k}$$

mentre i casi possibili sono tutte quelle combinazioni che contengono Mario e che sono

$$\binom{N-1}{k-1}$$

Facendo il rapporto si vede facilmente che $p = k/N$.

(b) Le settimane sono indipendenti; pertanto questa probabilità è

$$(1-p)^m.$$

(c) La probabilità è p.

(d) Il numero di settimane in cui Mario viene sorteggiato segue una legge binomiale con parametri l e p. La risposta quindi è

$$\sum_{j=s}^{l} \binom{l}{j} p^j (1-p)^{l-j} = 1 - \sum_{j=0}^{s-1} \binom{l}{j} p^j (1-p)^{l-j}$$

7. **Esercizio (13 giugno 2012 n. 1).** *Giuseppe si reca a scuola con l'autobus e deve trovarsi in aula alle 8.30. Egli lascia la propria abitazione tra le 8 e le 8.15 ed il tragitto con l'autobus, compreso il percorso fino alla fermata, prende dai 10 ai 20 minuti (a seconda del traffico, dei ritardi, siamo in Italia, etc.). Supponendo che il tempo di partenza dall'abitazione e la durata del tragitto siano indipendenti ed uniformemente distribuiti nel loro intervallo, qual'è la probabilità che Giuseppe arrivi a lezione in ritardo?*

Soluzione. Sia X il numero di minuti dopo le 8 in cui Giuseppe parte da casa. La distribuzione è uniforme sull'intervallo $[0, 15]$, per cui

$$f_X(x) = \frac{1}{15}, \qquad 0 \le x \le 15$$

Sia poi Y la durata del viaggio in autobus, distribuita uniformemente su $[10, 20]$ e quindi

$$f_Y(y) = \frac{1}{10}, \qquad 10 \le y \le 20$$

Le due variabili X ed Y sono indipendenti, per cui la distribuzione congiunta è

$$f(x,y) = \frac{1}{150}, \qquad 0 \le x \le 15, \quad 10 \le y \le 20$$

Giuseppe arriva in ritardo quando $X + Y > 30$, quindi la domanda è $P(X + Y > 30)$ ed è data da

$$P(X+Y>30) = \int\int_D f(x,y)\,dx\,dy$$

dove D è il dominio dove $f(x,y) \neq 0$ e $x + y > 30$. La retta $y = 30 - x$ interseca la retta $y = 20$ nel punto soluzione del sistema

$$\begin{aligned} y &= 30 - x \\ y &= 20 \end{aligned}$$

che è $(x = 10, y = 20)$. Pertanto

$$P(X+Y>30) = \int_{10}^{15} dx \int_{30-x}^{20} \frac{1}{150}\,dy = \frac{1}{12}$$

8. **Esercizio (31 marzo 2012 n. 1).** *Al pronto soccorso di un ospedale si presentano mediamente 5 pazienti all'ora, distribuiti secondo la legge di Poisson. Inoltre, la gravità dei casi viene catalogata con quattro codici: rosso, giallo, verde e bianco (in ordine decrescente di gravità). Il codice rosso si presenta n_R volte su 100, il codice giallo metà volte del rosso, il verde ed il bianco (in proporzioni uguali) metà volte del giallo. Qual'è la probabilità che, in un dato intervallo di un'ora, si presentino più di due persone da codice rosso?*

Soluzione. Sia X la v.a. che indica il numero di pazienti che arrivano al pronto soccorso in un'ora. Essendo distribuita secondo la legge di Poisson abbiamo che

$$P(X = k) = e^{-\lambda}\,\frac{\lambda^k}{k!}.$$

Siano ora X_R, X_G, X_V ed X_B le v.a. che indicano il numeri di pazienti in arrivo con codici rispettivamente rosso, giallo, verde e bianco. Siano inoltre n_R, n_G, n_V ed n_B le frazioni di codice rosso, giallo, verde e bianco su 100 pazienti. Dai dati del problema abbiamo che

$$\begin{aligned} & n_R + n_G + n_V + n_B = 100 \\ & n_G = n_R/2 \\ & n_V = n_B = n_G/2 \end{aligned}$$

La soluzione del sistema è immediata ed offre $n_R = 50$, $n_G = 25$, $n_V = n_B = 12.5$. Le variabili X_R, X_G, X_V ed X_B sono pertanto distribuite secondo la legge di Poisson con $\lambda_R = \lambda/2$, $\lambda_G = \lambda/4$ e $\lambda_V = \lambda_B = \lambda/8$. La variabile di interesse è X_R, per la quale abbiamo

$$P(X_R = k) = e^{-\lambda_R}\,\frac{\lambda_R^k}{k!}.$$

La risposta alla domanda è dunque

$$P(X_R > 2) = \sum_{k=3}^{\infty} e^{-\lambda_R}\,\frac{\lambda_R^k}{k!} = 1 - \sum_{k=0}^{2} e^{-\lambda_R}\,\frac{\lambda_R^k}{k!} = 0.456$$

9. **Esercizio (2 febbraio 2012 n. 2).** *Il punteggio ottenuto dagli studenti in un compito in classe ad una scuola superiore segue una legge normale $N(\mu, \sigma^2)$ con $\mu = 7.5$ e $\sigma^2 = 4$.*

   (i) Qual'è la probabilità che uno studente prenda un voto inferiore a 4?

   (ii) Se al compito partecipano 30 studenti, qual'è la probabilità che ci sia un voto inferiore a 4?

   Soluzione.

   (i) Sia X il voto di uno studente a caso. Allora $X \sim N(\mu, \sigma^2)$ e dobbiamo calcolare

$$p = P(X < 4) = P\left(\frac{X-\mu}{\sigma} < \frac{4-7.5}{2}\right) = P(Z < -1.75) = \Phi(-1.75) = 1-\Phi(1.75) = 0.04$$

   (ii) Sia Y il numero di studenti con voto inferiore a 4. La variabile Y segue una distribuzione binomiale $B(30, p)$, quindi la probabilità richiesta è

$$P(Y = 1) = \binom{30}{1} p\,(1-p)^2 9 = 0.367$$

10. **Esercizio (7 giugno 2011).** *Si fa un controllo sul contenuto in peso di una certa medicina. Il valore fornito dalla ditta farmaceutica produttrice è di 125 mg di sostanza (principio attivo) per pastiglia. Un gruppo di venti misurazioni fornisce i seguenti risultati:*

    127.87, 124.42, 128.06, 123.43, 127.35, 128.52, 125.65, 127.71, 123.32, 123.26,
    128.35, 127.55, 124.4, 123.77, 126.85, 125.53, 127.39, 126.96, 123.15, 124.49

    Stabilire se questi dati confermano con un livello di significatività del 10%, del 5% e dell'1% il valore fornito dalla ditta.

    Soluzione. L'ipotesi nulla H_0 è $\mu = \mu_0 = 125$. Dai dati abbiamo che $n = 20$, $\overline{X}_n = 125.9$, $\sigma^2 = 3.8$ e $\sigma = 1.95$. I quantili della legge di Student con 19 gradi di libertà, $t_{\alpha/2}(19)$, sono: $t_{0.01/2}(19) = 2.861$, $t_{0.05/2}(19) = 2.093$ e $t_{0.1/2}(19) = 1.729$. Le regioni critiche per il test sono pertanto le parti esterne agli intervalli

$$\begin{aligned}
&[\mu_0 - \sigma\, 2.861/\sqrt{n}, \mu_0 + \sigma\, 2.861/\sqrt{n}] = [123.753, 126.247], \quad \text{al} \quad 1\% \\
&[\mu_0 - \sigma\, 2.093/\sqrt{n}, \mu_0 + \sigma\, 2.093/\sqrt{n}] = [124.088, 125.912], \quad \text{al} \quad 5\% \\
&[\mu_0 - \sigma\, 1.729/\sqrt{n}, \mu_0 + \sigma\, 1.729/\sqrt{n}] = [124.246, 125.754], \quad \text{al} \quad 10\%
\end{aligned}$$

11. **Esercizio (Ross (II Ed.) Cap. 8 n. 16).** *Un medico ritiene la temperatura basale media degli individui apparentemente sani sia cresciuta nel tempo e che sia ora maggiore di 37 ℃ e, per dimostrarlo, ha scelto a caso un campione di 100 individui sani. Se la temperatura media del campione è* $\overline{X}_n = 37.078$ *℃ e la deviazione standard campionaria è 0.61 ℃, questo dimostra l'ipotesi al 5 % di fiducia? Ed all' 1% ? Ed al 10% ?*

Soluzione. Supponendo una distribuzione normale, l'ipotesi nulla è $H_0 : \mu = \mu_0 = 37$, da testare contro l'ipotesi alternativa $H_1 : \mu > \mu_0$ in un test unilaterale (ad una coda). La regione critica del test, nei termini della variabile non standardizzata X, è $\overline{X}_n > \mu_0 + \delta_\alpha$. La dimensione del campione è grande abbastanza per giustificare l'uso della distribuzione normale. Dalle tavole della distribuzione normale, individuiamo i quantili da usare: 1.282 al 10 %, 1.645 al 5 % e 2.326 all' 1%. I valori di δ_α sono rispettivamente 0.078, 0.1 e 0.142 che danno i valori 37.078, 37.100 e 37.142 per la regione critica. Il valore della media campionaria è inferiore a tutti e tre i valori e pertanto si accetta H_0 a tutti e tre i livelli di fiducia.

12. **Esercizio (Ross (II Ed.) Cap. 8 n. 23).** *Un produttore di condensatori afferma che la tensione di breakdown di un certo modello è mediamente superiore a 100 volt. Un campione di 12 trasformatori di questo modello fornisce i seguenti valori per la tensione di breakdown: 96, 98, 105, 92, 111, 114, 99, 103, 95, 101, 106, 97. Si può dire, all'1, al 5 ed al 10 % del livello di confidenza, che i dati confermano l'affermazione del costruttore?*

Soluzione. Supponendo una distribuzione normale, l'ipotesi nulla è $H_0 : \mu = \mu_0 = 100$, da testare contro l'ipotesi alternativa $H_1 : \mu < \mu_0$ in un test unilaterale (ad una coda). La media campionaria è

$$\overline{X}_n = 101.417$$

e la varianza campionaria

$$S^2 = 43.901, \quad S = 6.626.$$

La regione critica del test, nei termini della variabile non standardizzata X, è $\overline{X}_n < \mu_0 - \delta_\alpha$. La dimensione del campione non è grande abbastanza per giustificare l'uso della distribuzione normale e dobbiamo usare i quantili di Student. Dalle tavole della distribuzione di Student, individuiamo i quantili da usare: 1.363 al 10 %, 1.796 al 5 % e 2.718 all' 1%. I valori di δ_α sono rispettivamente 2.607, 3.436 e 5.199 che danno i valori 97.393, 96.565 e 94.801 per la regione critica. Il valore della media campionaria è superiore a tutti e tre i valori e pertanto si accetta H_0 a tutti e tre i livelli di fiducia.

13. **Esercizio (Ross (II Ed.) Cap. 8 n. 25).** *Uno scienziato che si occupa di inquinamento ambientale vuole verificare se due campioni di soluzioni in suo possesso possono provenire dalla stessa sorgente. Se fosse così, i pH delle due soluzioni dovrebbero coincidere e, per stabilire se questo sia vero, vengono fatte 10 misurazioni indipendenti per ciascuna soluzione. Il metodo usato garantisce che i valori misurati hanno una distribuzione normale con media pari al pH vero e deviazione standard* $\sigma = 0.05$. *I dati sono i seguenti: Verificare se i dati confermano*

Soluzione A	6.24	6.31	6.28	6.30	6.25	6.26	6.24	6.29	6.22	6.28
Soluzione B	6.27	6.25	6.33	6.27	6.24	6.31	6.28	6.29	6.34	6.27

o meno l'uguaglianza dei valori medi del pH delle due soluzioni ai livelli dell'1, 5 e 10 % di significatività.

Soluzione. L'ipotesi nulla H_0 è che $\mu_A = \mu_B$ ed il test è bilaterale a varianza nota, con $H_1 : \mu_A \neq \mu_B$. Le due medie campionarie sono

$$\overline{X}_A = 6.267 \qquad \overline{X}_B = 6.285$$

ed abbiamo $n = 10$ per il rango del campione. I quantili della distribuzione normale da utilizzare sono 1.282 al 10 %, 1.645 al 5 % e 2.326 all' 1%. Per la variabile standardizzata che rappresenta la differenza delle due medie,

$$Z = \frac{|\mu_A - \mu_B|}{\sqrt{2\,\sigma^2/n}}$$

otteniamo il valore $Z = 0.805$, che è chiaramente inferiore ai quantili di tutti e tre i livelli, e quindi i dati del campione non permettono di rifiutare l'ipotesi nulla.

14. **Esercizio (Ross (II Ed.) Cap. 11 n. 1).** *Secondo la teoria genetica di Mendel, una certa pianta di piselli dovrebbe produrre fiori bianchi, rosa o rossi con probabilità rispettivamente di* $p_1 = 1/4$, $p_2 = 1/2$ *e* $p_3 = 1/4$. *Per verificare la teoria, si studia un campione di* $n = 564$ *piante di cui 141 produssero fiori bianchi, 291 fiori rosa e 132 fiori rossi. Verificare la correttezza delle ipotesi mediante il test del* χ^2 *al 5 % di significatività.*

Soluzione. Il numero di classi è $k = 3$ e le frequenze osservate sono $x_1 = 141$, $x_2 = 291$ e $x_3 = 132$. Le frequenze attese sono $n/4$, $n/2$ ed $n/4$. La statistica T è data da

$$T = \sum_{i=1}^{3} \frac{(x_i - n\,p_i)^2}{n\,p_i} = 0.8617$$

Il numero di gradi di libertà è $l = k - 1$ ed il valore dato dalle tavole è di $\chi^2_{0.05}(2) = 5.991$. Il test non permette di rifiutare l'ipotesi nulla.

15. **Esercizio (Ross (II Ed.) Cap. 11 n. 4).** *Si pensa che il numero di interruzioni quotidiane di energia elettrica in una data città abbia una distribuzione di Poisson di media 4.2. I dati su 150 giorni sono riportati nella tabella. Verificare l'ipotesi con il test del* χ^2 *al 5% di confidenza.*

Interruzioni	0	1	2	3	4	5	6	7	8	9	10	11
Numero di giorni	0	5	22	23	32	22	19	13	6	4	4	0

Soluzione. Abbiamo un campione di rango $n = 150$ ed un numero di classi $k = 12$. Le probabilità da assegnare a ciascun valore sono date dalla legge di Poisson con $\lambda = 4.2$, quindi

$$p_i = e^{-\lambda} \frac{\lambda^i}{i!}$$

e sono date da

$$p = \{0.0150, 0.0630, 0.132, 0.185, 0.194, 0.163, 0.114, 0.0686, 0.0360, 0.0168, 0.00706, 0.00269\}$$

Se indichiamo con x_i le frequenze osservate, la statistica T è data da

$$T = \sum_{i=1}^{k} \frac{(x_i - n\, p_i)^2}{n\, p_i} = 16.35$$

Il valore del χ^2 da cercare sulle tavole è quello a $l = k - 1 = 11$ gradi di libertà al 5%, vale a dire $\chi^2_{0.05}(11) = 19.675$ che è maggiore del valore della statistica T. I dati del campione non permettono quindi di respingere l'ipotesi nulla.

16. **Esercizio (Ross (II Ed.) Cap. 11 n. 2).** *Si eseguono 1000 lanci di un dado per stabilire se sia truccato. Le frequenze osservate sono, rispettivamente, 158, 172, 164, 181, 160, 165 per la faccia "1, la faccia "2, etc. Verificare se il dado è bilanciato con un test del χ^2 di adattamento al 5 %.*

Soluzione. Se il dato è bilanciato le facce sono equiprobabili e le frequenze attese sono tutte uguali a 1000/6. Le classi sono 6 ($k = 6$, quindi $l = k - 1 = 5$). Indicando con x_i le frequenze osservate e con $p_i = 1/6$ le probabilità, abbiamo per la ststistica T:

$$T = \sum_{i=1,k} \frac{(x_i - n\, p_i)^2}{n\, p_i} = 2.18$$

Il valore del χ^2 da confrontare è $\chi^2_{0.05}(5) = 11.07$, che è più grande della statistica T. Il campione non permette di rifiutare l'ipotesi nulla di dado equilibrato.

17. **Esercizio (Ross (II Ed.) Cap. 11 n. 15).** *Un campione aleatorio di 500 nuclei familiari degli Stati Uniti è stato classificato secondo il reddito (in migliaia di dollari) e secondo la provenienza geografica ottenendo i risultati riportati nella tabella. Usando il test del χ^2 di*

Reddito	**Sud**	**Nord**
0-10	42	53
10-20	55	90
20-30	47	88
≥ 30	36	89

indipendenza al 5%, determinare se si può affermare che esiste una dipendenza del reddito dalla provenienza geografica.

Soluzione. L'ipotesi nulla H_0 è quella dell'indipendenza del reddito dalla provenienza geografica. Il rango del campione è $n = 500$, con $r = 4$ (classi di reddito) ed $s = 2$ (regioni di provenienza geografica). Indicando con $N_{ij}, i = 1, 4, \quad j = 1, 2$ i dati della tabella, calcoliamo le stime per le probabilità delle due variabili

$$p_i = \sum_{j=1}^{s} N_{ij}/n, \quad i = 1, ..., r \qquad q_j = \sum_{i=1}^{r} N_{ij}/n, \quad j = 1, ..., s$$

ottenendo i valori $p = \{0.19, 0.29, 0.27, 0.25\}$ e $q = \{0.36, 0.64\}$. Il calcolo della statistica T offre

$$T = \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{(N_{ij} - n\, p_i\, q_j)^2}{n\, p_i\, q_j} = 4.21$$

Il numero di gradi di libertà è dato da $l = (r-1)\,(s-1) = 3$ ed il valore del χ^2 che si ottiene dalle tavole è di $\chi^2_{0.05}(3) = 7.815$, che non permette di rifiutare l'ipotesi nulla.

18. **Esercizio (Ross (II Ed.) Cap. 11 n. 19).** *Si raccolgono i dati per verificare se l'ipertensione ed il fumo siano indipedenti, ottenendo i risultati riportati nella tabella. Usando il test*

	Non fumatori	**Fumatori moderati**	**Grandi fumatori**
Ipertesi	20	38	28
Non ipertesi	50	27	18

del χ^2 di indipendenza al 5%, determinare se si può affermare che esiste una dipendenza della tendenza all'ipertensione dal fumo.

Soluzione. L'ipotesi nulla H_0 è quella dell'indipendenza dell'ipertensione dal fumo. Il rango del campione è $n = 181$, con $r = 2$ (ipertensione) ed $s = 3$ (fumatori). Indicando con $N_{ij}, i = 1, 2, \quad j = 1, 3$ i dati della tabella, calcoliamo le stime per le probabilità delle due variabili

$$p_i = \sum_{j=1}^{s} N_{ij}/n, \quad i = 1, ..., r \qquad q_j = \sum_{i=1}^{r} N_{ij}/n, \quad j = 1, ..., s$$

ottenendo i valori $p = \{0.475, 0.525\}$ e $q = \{0.387, 0.359, 0.254\}$. Il calcolo della statistica T offre

$$T = \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{(N_{ij} - n\, p_i\, q_j)^2}{n\, p_i\, q_j} = 16.486$$

Il numero di gradi di libertà è dato da $l = (r-1)\,(s-1) = 2$ ed il valore del χ^2 che si ottiene dalle tavole è di $\chi^2_{0.05}(2) = 5.991$, che non permette di accettare l'ipotesi nulla.

19. **Esercizio (15 luglio 2014 n. 2).** *Il punteggio ottenuto dagli studenti in un compito in classe ad una scuola superiore segue una legge normale $N(\mu, \sigma^2)$ con $\mu = 7.5$ e $\sigma^2 = 4$.*

    (i) Qual'è la probabilità che uno studente prenda un voto inferiore a 4?

    (ii) (*) *Se al compito partecipano 30 studenti, qual'è la probabilità che ci sia almeno un voto inferiore a 4?*

    Soluzione.

    (i) Sia X il voto di uno studente preso a caso. Si chiede $P(X < 4)$. Standardizzando abbiamo

    $$\begin{aligned} P(X < 4) &= P\left(\frac{X - 7.5}{2} < \frac{4 - 7.5}{2}\right) = P\left(Z < \frac{4 - 7.5}{2}\right) = \\ &= P(Z < -1.75) = 1 - \Phi(1.75) = 0.04 \end{aligned}$$

    (ii) Sia Y il numero di voti inferiori a 4 su 30 studenti. Allora $Y \sim B(30, 0.04)$ e la risposta è $1 - P(Y = 0) = 0.71$.

20. **Esercizio (Lebfevre Cap. 3 Q. 28 p. 97).** *Una certa apparecchiatura ha bisogno di 59 transistor non difettosi per funzionare. Abbiamo a disposizione 60 pezzi fabbricati da una macchina che produce il 5% di transistor difettosi.*

    (i) Calcolare la probabilità che l'apparecchiatura funzioni.

    (ii) Calcolare la stessa probabilità usando la legge di Poisson.

    Soluzione.

(i) Il numero di transistor difettosi X segue la distribuzione binomiale $B(n, p)$ con $n = 60$ e $p = 0.05$. Si chiede che non ci sia più di un transistor difettoso sui 60 a disposizione, pertanto

$$\begin{aligned} P(X \leq 1) &= P(X = 0) + P(X = 1) = \binom{60}{0} p^0\,(1-p)^{60} + \binom{60}{1} p\,(1-p)^{59} = \\ &= (0.95)^{60} + 3\,(0.95)^{59} \approx 0.1916 \end{aligned}$$

(ii) Abbiamo $\lambda = n\,p = 3$ e quindi

$$P(X \leq 1) = P(X = 0) + P(X = 1) = e^{-\lambda}\left(\frac{\lambda^0}{0!} + \frac{\lambda^1}{1!}\right) \approx 1.991$$

21. **Esercizio (Lebfevre Cap. 3 Q. 46 p. 108).** *Il numero di particelle emesse da una sorgente radioattiva è descritto da una v.a. X distribuita secondo la legge di Poisson con media $\lambda = \ln 5$ all'ora. Da un'ora all'altra, inoltre, le emissioni sono indipendenti.*

    (i) Calcolare la probabilità che in almeno 30 periodi di un'ora, sulle 168 ore di una settimana, non ci siano state emissioni di particelle.

    (ii) Calcolare la stessa probabilità usando un'appropriata legge di Poisson.

    Soluzione.

    (i) La probabilità che in un'ora non ci siano emissioni è data dalla legge di Poisson per $k = 0$, $P(X = 0) = e^{-\lambda} = 1/5$. Quindi, indicando con X il numero di periodi di un'ora in cui non ci sono emissioni, abbiamo $X \sim B(n, p)$ con $n = 168$ e $p = 1/5 = 0.2$. La risposta pertanto è

    $$P(X \geq 30) = \sum_{k=30}^{168} \binom{168}{k} (0.2)^k\,(0.8)^{138-k} = 1 - \sum_{k=0}^{29} \binom{168}{k} (0.2)^k\,(0.8)^{138-k} \approx 0.78$$

    (ii) Si tratta di approssimare la distribuzione binomiale del punto precedente con Poisson usando $\lambda = n\,p = 168 \times 0.2 = 33.6$. Otteniamo:

    $$P(X \geq 30) = \sum_{k=30}^{\infty} e^{-\lambda}\,\frac{\lambda^k}{k!} = 1 - \sum_{k=0}^{29} e^{-\lambda}\,\frac{\lambda^k}{k!} \approx 0.76$$

22. **Esercizio (Lebfevre Cap. 4 Q. 30 p. 143).** *Il numero di inondazioni che avvengono in una certa regione ̀descritto da una v.a. che segue la legge di Poisson con media $\alpha = 2$ all'anno, indipendenti da anno ad anno. La durata delle inondazioni (in giorni), inoltre, segue una legge esponenziale con parametro $\lambda = 1/5$. Usando l'approssimazione normale, calcolare la probabilità che*

    (i) ci siano almento 80 inondazioni nella regione nei prossimi 50 anni;

    (ii) il tempo totale in cui il terreno risulta inondato nelle prossime 50 inondazioni sia inferiore a 200 giorni.

    Soluzione.

(i) Sia X_i il numero di inondazioni nell'$i-$esimo anno. La probabilità che dobbiamo calcolare è $P(X_1 + X_2 + ... + X_n \geq 80)$, dove $n = 50$. Usando l'approssimazione normale abbiamo

$$P(X_1 + X_2 + ... + X_n \geq 80) = P\left(\frac{X_1 + X_2 + ... + X_n - n\,\alpha}{\alpha\sqrt{n}} \geq \frac{80 - 100}{2\sqrt{50}}\right) =$$
$$P(Z \geq -\sqrt{2}) = 1 - \Phi(-\sqrt{2}) = \Phi(\sqrt{2}) \approx \Phi(1.41) \approx 0.9207$$

(ii) Sia Y_i la durata dell'$i-$esima inondazione. Dobbiamo calcolare $P(Y_1+Y_2+...+Y_n \leq 200)$, dove $n = 50$. Usando l'approssimazione normale abbiamo

$$P(Y_1 + Y_2 + ... + Y_n \leq 200) = P\left(\frac{Y_1 + Y_2 + ... + Y_n - n/\lambda}{\sqrt{n}/\lambda} \leq \frac{200 - 250}{5\sqrt{50}}\right) =$$
$$P(Z \leq -\sqrt{2}) = \Phi(-\sqrt{2}) \approx 1 - \Phi(1.41) \approx 1 - 0.9207 = 0.079$$

23. **Esercizio (20 settembre 2011 n. 2).** *Una fabbrica produce televisori al plasma; sia p la frazione (assoluta, non percentuale) di televisori difettosi. Esprimere la probabilità che, su un campione di N televisori, k siano difettosi, usando sia la distribuzione binomiale che quella di Poisson. Posto, quindi, $N = 120$ e $k = 3$, confrontare i due valori (binomiale e Poisson) per 10 valori di $p = 0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09$ e 0.1, commentando i risultati.*

Soluzione. Sia X il numero di televisori difettosi; la variabile segue la legge binomiale di parametri N e p, quindi

$$P(X = k) = \binom{N}{k} p^k\,(1-p)^{N-k}.$$

Usando Poisson, con $\lambda = N\,p$ abbiamo invece

$$P(X = k) = e^{-\lambda}\,\frac{\lambda^k}{k!}.$$

Usando i dati del problema con le due distribuzioni otteniamo la seguente tabella:

p	Prob. Binom.	prob. Poisson
0.01	0.0867	0.867
0.02	0.211	0.209
0.03	0.215	0.212
0.04	0.151	0.152
0.05	0.0869	0.0892
0.06	0.0435	0.0464
0.07	0.0198	0.0222
0.08	0.00833	0.00999
0.09	0.00330	0.00428
0.1	0.00124	0.00177

24. **Esercizio (Lebfevre Cap. 3 Q. 46 p. 108).** *Il numero di particelle emesse da una sorgente radioattiva è descritto da una v.a. X distribuita secondo la legge di Poisson con media $\lambda = \ln 5$ all'ora. Da un'ora all'altra, inoltre, le emissioni sono indipendenti.*

   (i) *Calcolare la probabilità che in almeno 30 periodi di un'ora, sulle 168 ore di una settimana, non ci siano state emissioni di particelle.*

   (ii) *Calcolare la stessa probabilità usando un'appropriata legge di Poisson.*

Soluzione.

(i) La probabilità che in un'ora non ci siano emissioni è data dalla legge di Poisson per $k = 0$, $P(X = 0) = e^{-\lambda} = 1/5$. Quindi, indicando con X il numero di periodi di un'ora in cui non ci sono emissioni, abbiamo $X \sim B(n, p)$ con $n = 168$ e $p = 1/5 = 0.2$. La risposta pertanto è

$$P(X \geq 30) = \sum_{k=30}^{168} \binom{168}{k} (0.2)^k \, (0.8)^{138-k} = 1 - \sum_{k=0}^{29} \binom{168}{k} (0.2)^k \, (0.8)^{138-k} \approx 0.78$$

(ii) Si tratta di approssimare la distribuzione binomiale del punto precedente con Poisson usando $\lambda = n\,p = 168 \times 0.2 = 33.6$. Otteniamo:

$$P(X \geq 30) = \sum_{k=30}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} = 1 - \sum_{k=0}^{29} e^{-\lambda} \frac{\lambda^k}{k!} \approx 0.76$$

25. Ogni confezione di biscotti di gusto nocciola contiene 100 biscotti. Il numero di pezzi di nocciole presenti in un singolo biscotto può essere modellato con una variabile aleatoria di Poisson di parametro 2. Si chiede:

   - Quanto vale la probabilità che un biscotto preso a caso da una confezione contenga più di un pezzo di nocciola?
   - Quanto vale la probabilità di trovare più di 2 pezzi di nocciole (in totale) in due biscotti presi a caso da una confezione?
   - Quanto vale la probabilità di trovare più di 160 pezzi di nocciole in tutta la confezione?

Soluzione.

- Sia X il numero di pezzi di nocciola contenuti in un biscotto preso a caso. La probabilità richiesta è

$$P(X > 1) = P(X \geq 2) = 1 - P(X \leq 1) = 1 - [P(X = 0) + P(X = 1)] =$$
$$= 1 - e^{-\lambda} \left(1 + \frac{\lambda}{1}\right) = 1 - e^{-2} \left(1 + \frac{2}{1}\right) = 1 - 3\,e^{-2} \approx 0.59$$

- Siano X_1 ed X_2 le variabili che indicano i pezzi di nocciola contenuti rispettivamente nel primo e nel secondo biscotto. La probabilità richiesta è $P(X_1+X_2>2)$ che è una variabile di Poisson di parametro $2\,\lambda=4$. Quindi, indicando con $Y=X_1+X_2$:

$$P(Y>2)=1-P(Y\leq 2)=1-[P(Y=0)+P(Y=1)+P(Y=2)]=$$
$$=1-e^{-4}\left(1+\frac{4}{1}+\frac{4^2}{2}\right)=1-13\,e^{-2}\approx 0.76$$

- Indicando con X_1 i pezzi di nocciola nel primo biscotto, con X_2 quelli nel secondo, e così via, la probabilità richiesta è $P(X_1+...+X_n>160)=1-P(X_1+...+X_n\leq 160)$, con $n=100$. Usando l'approssimazione normale con la correzione di continuità, e ricordando che per una variabile di Poisson X di parametro λ si ha $E[X]=Var(X)=\lambda$,

$$P(X_1+...+X_n\leq 160)\approx P(X_1+...+X_n\leq 159.5)=P\left(S_n\leq\frac{159.5-100\times 2}{\sqrt{100\times 2}}\right)=$$
$$=P\left(S_n\leq -2.86\right)=1-P\left(S_n\leq 2.86\right)=1-0.9979$$

e quindi $P(X_1+...+X_n>160)\approx 0.9979$.

Il calcolo esatto non è fattibile con una normale calcolatrice; notando che la variabile S_n è una variabile di Poisson di parametro 200 ed usando un software appropriato otteniamo

$$P(S_n>160)=1-P(S_n\leq 160)=1-\sum_{k=0}^{160}e^{-200}\,\frac{200^k}{k!}=0.9980$$

26. Un automobilista entra in un parcheggio e vuole parcheggiare la sua automobile in un ben preciso posto-auto. Davanti a lui ci sono 5 macchine, ciascuna con probabilità $p=0.5$ di occupare quel posto. Qualè la probabilità che la macchina davanti a lui lo occupi?

Soluzione. $P(X=5)=(0.8)^4\,(0.2)=0.082$

27. In un controllo a campione sulle emissioni la probabilità di trovare un'automobile fuori norma è di 0.1 %. Qualè la probabilità che si debbano effettuare più di 3 controlli prima di trovarne una a norma?

Soluzione.

$$P(X>3)=1-P(X\leq 3)=1-\sum_{k=1}^{3}P(X=k)=1-\sum_{k=1}^{3}(0.1)^{k-1}(0.9)=0.001$$

28. Uno zoo ha a disposizione 100 kg di carne per sfamare un leone. La quantità consumata dal leone in un giorno è rappresentata da una variabile casuale di media $\mu = 5$ kg e deviazione standard $\sigma = 1$ kg.

   (i) Usando l'approssimazione normale, calcolare la probabilità che la carne a disposizione basti per 21 giorni;

   (ii) supponiamo che in 10 giorni il leone abbia mangiato 60 kg; supponendo sempre $\sigma = 0.5$ kg, determinare l'intervallo di confidenza per la media al 99%;

   (iii) possiamo ancora giustificare che la media sia $\mu = 5$ kg ?

Soluzione. Sia $X \sim \mathcal{N}(\mu, \sigma^2)$, con $\mu = 5$ e $\sigma = 1$. Siano $X_1, X_2, ..., X_n$ (con $n = 21$) le variabili he rappresentano il consumo di carne (da parte del leone) nel primo, nel secondo, ..., nel 21-esimo giorno. La carne a disposizione basta se $X_1 + X_2 + ... + X_n \leq 100$. Usando l'approssimazione normale abbiamo che

$$P(X_1 + X_2 + ... + X_n \leq 100) = P\left(\frac{X_1 + X_2 + ... + X_n - n\,\mu}{\sigma\sqrt{n}} \leq \frac{100 - 21 \times 5}{1 \times \sqrt{21}}\right) =$$
$$P(S_n \leq -1.09) \approx \Phi(-1.09) = 1 - \Phi(1.09) \approx 1 - 0.8621 = 0.01379$$

29. La dose di caffè necessaria per una tazzina può essere rappresentata da una variabile casuale X di media $\mu = 7$ gr e scarto quadratico medio $\sigma = 0.5$ gr.

   (i) Usando l'approssimazione normale, calcolare la probabilità che un barattolo da 250 gr sia sufficiente per 36 tazzine;

   (ii) supponiamo che per le prime 20 tazzine siano stati usati 140 gr; supponendo sempre $\sigma = 0.5$ gr, determinare l'intervallo di confidenza per la dose media al 99%;

   (iii) possiamo ancora giustificare che la dose media sia $\mu = 7$ gr ?

Soluzione. Sia $X \sim \mathcal{N}(\mu, \sigma^2)$, con $\mu = 7$ e $\sigma = 0.5$. Siano $X_1, X_2, ..., X_n$ (con $n = 36$) le variabili che rappresentano le dosi di caffè della prima, della seconda, ..., della 36-esima tazzina. Il barattolo di caffè a disposizione basta se $X_1 + X_2 + ... + X_n \leq 250$. Usando l'approssimazione normale abbiamo che

$$P(X_1 + X_2 + ... + X_n \leq 250) = P\left(\frac{X_1 + X_2 + ... + X_n - n\,\mu}{\sigma\sqrt{n}} \leq \frac{250 - 36 \times 7}{0.5 \times \sqrt{36}}\right) =$$
$$P(S_n \leq -0.67) \approx \Phi(-0.67) = 1 - \Phi(0.67) \approx 1 - 0.7486 = 0.2514$$

30. L'ente che gestisce un tratto di autostrada ha sale sufficiente per eliminare un accumulo di 2 metri di neve su tale tratto. Supponiamo che la quantità di neve che cade in un giorno sia rappresentata da una variabile casuale X di media $\mu = 4$ cm e deviazione standard $\sigma = 0.8$ cm.

(i) Usando l'approssimazione normale, calcolare la probabilità che il sale a disposizione basti per 50 giorni;

(ii) supponiamo che nei primi 10 giorni del periodo siano caduti 50 cm di neve; supponendo sempre $\sigma = 0.8$ cm, determinare l'intervallo di confidenza per la media al 99%;

(iii) possiamo ancora giustificare che la media sia $\mu = 4$ cm ?

Soluzione. Sia $X \sim \mathcal{N}(\mu, \sigma^2)$, con $\mu = 4$ e $\sigma = 0.8$. Siano $X_1, X_2, ..., X_n$ (con $n = 50$) le variabili che rappresentano la quantità di neve della prima, della seconda, ..., della 50-esima giornata. La quantità di sale a disposizione basta se $X_1 + X_2 + ... + X_n \leq 200$. Usando l'approssimazione normale abbiamo che

$$P(X_1 + X_2 + ... + X_n \leq 200) = P\left(\frac{X_1 + X_2 + ... + X_n - n\,\mu}{\sigma\,\sqrt{n}} \leq \frac{200 - 50 \times 4}{0.8 \times \sqrt{50}}\right) =$$
$$P(S_n \leq 0) \approx \Phi(0) = 0.5$$

31. Una cartoleria ha una scorta di matite sufficiente per 400 persone. Supponiamo che il numero di persone che si presentano in un giorno sia una variabile casuale X di media $\mu = 18$ persone e deviazione standard $\sigma = 2$ persone.

(i) Usando l'approssimazione normale, calcolare la probabilità che il numero di matite a disposizione basti per 22 giorni;

(ii) supponiamo che nei primi 10 giorni del periodo si siano presentate 240 persone; supponendo sempre $\sigma = 2$, determinare l'intervallo di confidenza per la media giornaliera al 99%;

(iii) possiamo ancora giustificare che la media sia $\mu = 18$ persone ?

Soluzione. Sia $X \sim \mathcal{N}(\mu, \sigma^2)$, con $\mu = 18$ e $\sigma = 2$. Siano $X_1, X_2, ..., X_n$ (con $n = 22$) le variabili che rappresentano il numero di clienti della prima, della seconda, ..., della 22-esima giornata. La scorta di matite a disposizione basta se $X_1 + X_2 + ... + X_n \leq 400$. Usando l'approssimazione normale abbiamo che

$$P(X_1 + X_2 + ... + X_n \leq 400) = P\left(\frac{X_1 + X_2 + ... + X_n - n\,\mu}{\sigma\,\sqrt{n}} \leq \frac{400 - 22 \times 18}{2 \times \sqrt{22}}\right) =$$
$$P(S_n \leq 0.43) \approx \Phi(0.43) \approx 0.6664$$

32. Ad un turno elettorale, gli scrutini iniziali eseguiti su un numero di sole 20 sezioni danno per il partito A le seguenti percentuali:

35, 60, 55, 40, 49, 52, 39, 46, 55, 45, 70, 10, 22, 34, 44, 62, 58, 57, 43, 20

Determinare gli intervalli di confidenza per il risultato finale del partito A al 90%, 95% e 99%.

Soluzione. Il rango del campione è $n = 20$ e dobbiamo usare la distribuzione di Student con 19 gradi di libertà. Dalle tavole di Student abbiamo: $t_{0.05}(19) = 1.729$, $t_{0.025}(19) = 2.093$ e $t_{0.005}(19) = 2.861$. Dai dati abbiamo $\overline{X} = 44.8$ e $S^2 = 230.7$. Gli intervalli di confidenza sono pertanto:

$$\text{Al 90\%:} \quad \left[\overline{X} - \frac{S}{\sqrt{n}}\, t_{0.05}(19), \quad \overline{X} + \frac{S}{\sqrt{n}}\, t_{0.05}(19)\right] = [38.93, 50.67]$$

$$\text{Al 95\%:} \quad \left[\overline{X} - \frac{S}{\sqrt{n}}\, t_{0.025}(19), \quad \overline{X} + \frac{S}{\sqrt{n}}\, t_{0.025}(19)\right] = [37.69, 51.91]$$

$$\text{Al 99\%:} \quad \left[\overline{X} - \frac{S}{\sqrt{n}}\, t_{0.005}(19), \quad \overline{X} + \frac{S}{\sqrt{n}}\, t_{0.005}(19)\right] = [35.08, 54.52]$$

33. Un insieme di misurazioni del punto di fusione del piombo fornisce i seguenti dati, in gradi centigradi:

330, 328.6 342.4 334 337.5 341 343.3 329.5 322 331

Supponendo che i dati provengano da una popolazione normale di varianza incognita, determinare gli intervalli di confidenza al 90%, 95 % e 99% per la media.

Soluzione. Il rango del campione è $n = 10$ e dobbiamo usare la distribuzione di Student con 9 gradi di libertà. Dalle tavole di Student abbiamo: $t_{0.05}(9) = 1.833$, $t_{0.025}(9) = 2.262$ e $t_{0.005}(9) = 3.25$. Dai dati abbiamo $\overline{X} = 333.9$ e $S^2 = 48.5$. Gli intervalli di confidenza sono pertanto:

$$\text{Al 90\%:} \quad \left[\overline{X} - \frac{S}{\sqrt{n}}\, t_{0.05}(9), \quad \overline{X} + \frac{S}{\sqrt{n}}\, t_{0.05}(9)\right] = [329.9, 338.0]$$

$$\text{Al 95\%:} \quad \left[\overline{X} - \frac{S}{\sqrt{n}}\, t_{0.025}(9), \quad \overline{X} + \frac{S}{\sqrt{n}}\, t_{0.025}(9)\right] = [328.9, 338.9]$$

$$\text{Al 99\%:} \quad \left[\overline{X} - \frac{S}{\sqrt{n}}\, t_{0.005}(9), \quad \overline{X} + \frac{S}{\sqrt{n}}\, t_{0.005}(9)\right] = [326.8, 341.1]$$

34. **(Ross 7.38)** Un insieme di misurazioni della capacità di 10 batterie fornisce i seguenti risultati, in ampere-ora:

140, 136 150 144 148 152 138 141 143 151

Stimare la varianza della popolazione e fornirne gli intervalli di confidenza al 90%, 95 % e 99%.

Soluzione. Il rango del campione è $n = 10$ e dobbiamo usare la distribuzione del χ^2 con 9 gradi di libertà. I valori di α che ci interessano sono 0.05 e 0.95 per l'intervallo al 90%, 0.025 e 0.975 per l'intervallo al 95%, 0.005 e 0.995 per l'intervallo al 99%, Dalle tavole del χ^2 abbiamo: La varianza campionaria è $S^2 = 32.23$. Gli intervalli di confidenza sono dati dalla formula

Livello	$\chi^2_{\alpha/2}(9)$	$\chi^2_{1-\alpha/2}(9)$
90%	16.919	3.325
95%	19.023	2.700
99%	23.589	1.735

$$\left[\frac{(n-1)\,S^2}{\chi^2_{\alpha/2}(n-1)}, \frac{(n-1)\,S^2}{\chi^2_{1-\alpha/2}(n-1)}\right]$$

e pertanto:

$$\begin{aligned} &\text{Al } 90\%: && [17.14, 87.25] \\ &\text{Al } 95\%: && [15.25, 107.44] \\ &\text{Al } 99\%: && [12.30, 167.20] \end{aligned}$$

35. **Esercizio (12 aprile 2011).** *Da un'urna che contiene* 20 *palline bianche e* 30 *nere vengono estratte* 4 *palline con reimbussolamento. Qual è la probabilità che venga estratta per* 3 *(tre) volte una pallina bianca?*

Soluzione. La probabilità di estrarre una pallina bianca è 2/5. Il numero di palline bianche su 4 estrazioni è pertanto una variabile binomiale del tipo $B(4, 2/5)$ e la probabilità che ce ne siano 3 bianche su 4 è, indicando con X la variabile aleatoria in questione,

$$P(X = 3) = \binom{4}{3}\left(\frac{2}{5}\right)^3\left(\frac{3}{5}\right)^1 = \frac{96}{625} \approx 0.15$$

36. **Esercizio (21 giugno 2011).** *In un dado non equilibrato, la probabilità che si presenti la faccia "2" è tripla rispetto alla probabilità delle altre facce (che è uniforme). Usando sia il calcolo esatto che l'approssimazione normale, calcolare la probabilità che, su* 18 *lanci,*

   (i) il 2 *compaia un numero di volte compreso tra* 4 *e* 9*;*

(ii) il 2 non compaia mai.

Qual'è l'errore relativo commesso con l'approssimazione normale?

Soluzione. Si vede subito che $p = P(\text{esce la faccia "2"}) = 3/8$. Siano ora X_k, $k = 1, 2, ..., n$, con $n = 18$, variabili di Bernoulli così definite: $X_k = 1$ se esce la faccia 2 al $k-$esimo lancio e $X_k = 0$ altrimenti. Ovviamente, $P(X_k = 1) = p$. Le due domande del testo si possono allora tradurre con:

(i)

$$4 \leq \sum_{k=1}^{n} X_k \leq 9$$

(ii)

$$\sum_{k=1}^{n} X_k = 0$$

La risposta esatta alle due domande si trova notando che la variabile $X = \sum_{k=1}^{n} X_k$ è una variabile binomiale di parametri $n = 18$ e $p = 3/8$. Quindi:

(i)

$$P(4 \leq X \leq 9) = \sum_{k=4}^{9} \binom{n}{k} p^k \, (1-p)^{n-k} \approx 0.86$$

(ii)

$$X = 0.0002$$

Per usare l'approssimazione normale applichiamo la correzione di continuità, modificando le domande in:

(i) $3.5 < X < 9.5$;

(ii) $-0.5 < X < 0.5$.

(i)

$$\begin{aligned} P(4 \leq X \leq 9) &\approx P(3.5 \leq X \leq 9.5) = P\left(\frac{3.5 - n\,p}{\sigma\sqrt{n}} < \frac{X - n\,p}{\sigma\sqrt{n}} < \frac{9.5 - n\,p}{\sigma\sqrt{n}}\right) = \\ &= P\left(\frac{3.5 - 18\,(3/8)}{\sqrt{(3/8)\,(5/8)\,18}} < Z < \frac{9.5 - 18\,(3/8)}{\sqrt{(3/8)\,(5/8)\,18}}\right) = P(-1.58 < Z < 1.34) = \\ &= \Phi(1.34) - \Phi(-1.58) = \Phi(1.34) - 1 + \Phi(1.58) = 0.9099 + 0.9429 - 1 = 0.85 \end{aligned}$$

Errore relativo: $(0.86 - 0.85)/0.86 = 0.01$ cioè l'1 %.

(ii)

$$P(X=0) \approx P(-0.5 < X < 0.5) = P\left(\frac{-0.5 - n\,p}{\sigma\,\sqrt{n}} < \frac{X - n\,p}{\sigma\,\sqrt{n}} < \frac{0.5 - n\,p}{\sigma\,\sqrt{n}}\right) =$$
$$= P\left(\frac{-0.5 - 18\,(3/8)}{\sqrt{(3/8)\,(5/8)\,18}} < Z < \frac{0.5 - 18\,(3/8)}{\sqrt{(3/8)\,(5/8)\,18}}\right) = P(-3.53 < Z < -3.04) =$$
$$= \Phi(-3.04) - \Phi(-3.53) = 1 - \Phi(3.04) - 1 + \Phi(3.53) =$$
$$\Phi(3.53) - \Phi(3.04) = 0.9998 - 0.9988 = 0.001$$

Errore relativo: $(0.001 - 0.0002)/0.0002 = 4$, cioè il 400 %.

37. **Esercizio (12 novembre 2011).** *Una compagnia che produce materiale elettronico sostiene che nella sua produzione di un certo componente vi è una percentuale di pezzi difettosi solo del 2%. Usando l'approssimazione normale:*

(i) *Dare una stima della probabilità che in un lotto di 120 pezzi ve ne siano almeno 3 difettosi.*

(ii) *In un controllo di verifica, sono stati trovati 9 pezzi difettosi su un campione di 200. Cosa si può dire sulle affermazioni della compagnia? E se fossero stati trovati 16 pezzi difettosi?*

Soluzione.

(i) Sia X_k, $k = 1, 2, ..., n$, la variabile di Bernoulli che vale 1 se il $k-$esimo pezzo è difettoso e 0 se è funzionante, con $n = 120$. Il numero di pezzi difettosi X è dato dalla somma $X = X_1 + X_2 + ... + X_n$. Pertanto la probabilità richiesta è $P(X \geq 3)$. Siccome X è una variabile binomiale con $n = 120$ e $p = 0.02$ il calcolo esatto dà

$$P(X \geq 3) = 1 - P(X \leq 2) = 1 - \sum_{k=0}^{2} \binom{n}{k} p^k\,(1-p)^{n-k} = 0.431$$

mentre l'approssimazione di Poisson con $\lambda = n\,p$ offre

$$P(X \geq 3) = 1 - P(X \leq 2) = 1 - \sum_{k=0}^{2} e^{-\lambda}\,\frac{\lambda^k}{k!} = 0.430.$$

L'approssimazione normale, con $\sigma = p\,(1-p)$ e senza correzione di continuità, offre

$$P(X \geq 3) = 1 - P(X < 3) = 1 - P\left(\frac{X - n\,p}{\sigma\sqrt{n}} < \frac{3 - n\,p}{\sigma\sqrt{n}}\right) \approx$$
$$\approx 1 - P\,(Z < 0.39) = 1 - \Phi(0.39) = 1 - 0.6517 = 0.348$$

Utilizzando la correzione di continuità otteniamo invece

$$P(X \geq 2.5) = 1 - P(X < 2.5) = 1 - P\left(\frac{X - n\,p}{\sigma\sqrt{n}} < \frac{2.5 - n\,p}{\sigma\sqrt{n}}\right) \approx$$
$$\approx 1 - P\,(Z < 0.065) = 1 - \Phi(0.065) = 1 - 0.5279 = 0.472$$

(ii) In questo caso $n = 200$ e la percentuale campionaria di pezzi difettosi, $9/200 \approx 0.045$ è superiore alla media teorica $p = 0.02$. Abbiamo per la statistica del campione:

$$Z = \frac{\overline{X}_n - n\,p}{\sigma\,\sqrt{n}} = \frac{9 - 200\,(0.02)}{\sqrt{0.02(1-0.02)\,200}} \approx 4.3$$

da confrontare con i quantili z_α per $\alpha = 0.01$, 0.05 e 0.1 (è un test unilaterale), che sono rispettivamente 2.326, 1.645 e 1.282, che fanno rigettare l'affermazione della ditta.

38. **Esercizio (Ross (III Ed.) Cap. 8 n. 20).** *Una compagnia petrolifera dichiara che il contenuto di zolfo del suo carburante diesel non supera lo 0.15 %. Se ne analizzano 40 campioni, trovando un contenuto medio di 0.162 % con deviazione standard campionaria di 0.4 %. Possiamo confutare le affermazioni della compagnia al 5 % di significatività ?*

Soluzione. Con $\mu_0 = 0.15$ abbiamo

$$H_0 : \mu \le \mu_0 \qquad H_1 : \mu > \mu_0$$

Inoltre $n = 40$, $S_n = 0.4$ e $\overline{X}_n = 0.162$. Otteniamo quindi la statistica

$$Z = \frac{\overline{X}_n - \mu_0}{S_n}\,\sqrt{n} = \frac{0.162 - 0.15}{0.4}\,\sqrt{40} = 0.19$$

da confrontarsi con il quantile normale z_α per $\alpha = 0.05$, cioè $z_\alpha = 1.645$. Non si possono quindi confutare le affermazioni della compagnia.

39. **Esercizio (Ross (III Ed.) Cap. 8 n. 25).** *Uno scienziato che si occupa di inquinamento ambientale vuole capire se due campioni di soluzioni possono provenire dalla stessa sorgente. Se fosse così, i pH delle due soluzioni dovrebbero coincidere; vengono allora fatte 10 misurazioni indipendenti per ciascuna soluzione. Il metodo usato garantisce che i valori isurati hanno distribuzione normale con media pari al pH vero e deviazione standard di 0.05. I dati ottenuto sono i seguenti:*

Soluzione A:	$6.24, 6.31, 6.28, 6.30, 6.25, 6.26, 6.24, 6.29, 6.22, 6.28$
Soluzione B:	$6.27, 6.25, 6.33, 6.27, 6.24, 6.31, 6.28, 6.29, 6.34, 6.27$

(a) Tali dati mostrano una differenza nei pH al 5 % di significatività?

(b) Quanto vale il p-dei-dati di questo test?

Soluzione. Usiamo il test bilaterale per la differenza di due medie con varianza nota. Dai dati abbiamo $n_A = n_B = n = 10$, $\sigma = 0.05$; $\overline{X}_A = 6.267$, $\overline{X}_B = 6.285$. La statistica ci dà:

$$Z = \frac{|\overline{X}_A - \overline{X}_B|}{\sqrt{2\,\sigma^2}}\,\sqrt{n} = \frac{0.018}{\sqrt{2}\,0.05}\,\sqrt{10} = 0.805.$$

(a) Confrontiamo il valore di Z con il quantile normale $z_{\alpha/2}$ per $\alpha = 0.05$, cioè $z_{\alpha/2} = 1.960$: i dati non permettono di rigettare l'ipotesi nulla.

(b) Il p-dei-dati è

$$p = P(|Z| > 0.805) = 2\,P(Z > 0.805) = 2\,\{1 - P(Z < 0.805)\} \approx 2\,(1 - 0.7895) = 0.421$$

quindi l'ipotesi nulla viene accettata per ogni $\alpha < 0.421$, che è un valore molto elevato per α.

40. **Esercizio (Svolto in classe - 19/12/2016).** *Siano X e Y due variabili normali standard, cioè $X, Y \sim N(0,1)$. Calcolare la distanza media tra le due variabili aleatorie, vale a dire $E[|X-Y|]$.*

Soluzione. La risposta data in classe era giusta !! Cioè:

$$E[|X-Y|] = \frac{2}{\sqrt{\pi}}$$

Ci sono due vie. La prima è quella percorsa in classe. Abbiamo $W = X - Y \sim N(0,2)$ (cioè $E[W] = 0$ e $Var(W) = 2$) e quindi

$$f_W = \frac{1}{\sqrt{2}\,\sqrt{2\pi}}\,e^{-w^2/4} = \frac{1}{2\,\sqrt{\pi}}\,e^{-w^2/4}.$$

Pertanto

$$\begin{aligned} E[|W|] &= \int_{-\infty}^{\infty} |w|\,\frac{1}{2\,\sqrt{\pi}}\,e^{-w^2/4}\,dw = \frac{1}{\sqrt{\pi}}\int_0^{\infty} w\,e^{-w^2/4}\,dw = \\ &= \frac{1}{\sqrt{\pi}}\int_0^{\infty}\frac{d}{dw}\left(-2\,e^{-w^2/4}\right)dw = -\frac{2}{\sqrt{\pi}}\left[e^{-w^2/4}\right]_0^{\infty} = \frac{2}{\sqrt{\pi}} \end{aligned}$$

L'altro modo di procedere è quello di calcolare $E[|X-Y|]$ direttamente, cioè

$$\begin{aligned} E[|X-Y|] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}|x-y|\,f_{XY}(x,y)\,dx\,dy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}|x-y|\,f_X(x)\,f_Y(y)\,dx\,dy = \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}|x-y|\,\frac{1}{2\,\pi}\,e^{-(x^2+y^2)/2}\,dx\,dy = \frac{1}{2\,\pi}\int_{-\infty}^{\infty}e^{-x^2/2}\left[\int_{-\infty}^{\infty}|x-y|\,e^{-y^2/2}\,dy\right]dx = \\ &= \frac{1}{2\,\pi}\int_{-\infty}^{\infty}e^{-x^2/2}\left[\int_{-\infty}^{x}(x-y)\,e^{-y^2/2}\,dy + \int_x^{\infty}(y-x)\,e^{-y^2/2}\,dy\right]dx = \\ &= \frac{1}{2\,\pi}\int_{-\infty}^{\infty}e^{-x^2/2}x\,\left[\int_{-\infty}^{x}e^{-y^2/2}\,dy - \int_x^{\infty}e^{-y^2/2}\,dy\right]dx - \\ &\quad -\frac{1}{2\,\pi}\int_{-\infty}^{\infty}e^{-x^2/2}\left[\int_{-\infty}^{x}y\,e^{-y^2/2}\,dy - \int_x^{\infty}y\,e^{-y^2/2}\,dy\right]dx = \end{aligned}$$

... integrando per parti il primo termine ...

$$\begin{aligned} &= \frac{1}{2\,\pi}\left[-e^{-x^2/2}\left(\int_{-\infty}^{x}e^{-y^2/2}\,dy - \int_x^{\infty}e^{-y^2/2}\,dy\right)\right]_{-\infty}^{\infty} + \frac{1}{2\,\pi}\int_{-\infty}^{\infty}e^{-x^2/2}\left(e^{-x^2/2}+e^{-x^2/2}\right) - \\ &\quad -\frac{1}{2\,\pi}\int_{-\infty}^{\infty}e^{-x^2/2}\left(-e^{-x^2/2}-e^{-x^2/2}\right)dx = 2\,\frac{1}{\pi}\int_{-\infty}^{\infty}e^{-x^2} = \frac{1}{\pi}\sqrt{\pi} = \frac{2}{\sqrt{\pi}} \end{aligned}$$